SEQUENCE LISTING

<110> Covacci, Antonello
Bugnoli, Massimo
Telford, John
Macchia, Giovanni
Rappuoli, Rino

<120> Helicobacter Pylori CAI Antigen Proteins Useful For Vaccines And Diagnostics

<130> CHIR0157

<140> 09/360,685
<141> 1999-07-26

<160> 8

<170> PatentIn Ver. 2.1

<210> 1
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 1
gcaagcttat cgatgtcgac tcgagct 27

<210> 2
<211> 3960
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 2
aaaaagaaag gaagaaaatg gaaatacaac aaacacaccg caaaatcaat cgccctctgg 60
tttctctcgc tttagtagga gcattagtca gcatcacacc gcaacaaagt catgccgcct 120
ttttcacaac cgtgatcatt ccagccattg ttggggtat cgctacaggc accgctgtag 180
gaacggtctc agggcttctt agctgggggc tcaaacaagc cgaagaagcc aataaaaccc 240
cagataaacc cgataaagtt tggcgcattc aagcaggaaa aggctttaat gaattcccta 300

acaaggaata cgacttatac agatcccttt tatccagtaa gattgatgga ggttgggatt 360
gggggaatgc cgctaggcat tattgggtca aaggcgggca acagaataag cttgaagtgg 420
atatgaaaga cgctgtaggg acttatacct tatcagggct tagaaacttt actggtgggg 480
atttagatgt caatatgcaa aaagccactt tacgcttggg ccaattcaat ggcaattctt 540
ttacaagcta taaggatagt gctgatcgca ccacgagagt ggatttcaac gctaaaaata 600
tctcaattga taattttgta gaaatcaaca atcgtgtggg ttctggagcc gggaggaaag 660
ccagctctac ggttttgact ttgcaagctt cagaagggat cactagcgat aaaaacgctg 720
aaatttctct ttatgatggt gccacgctca atttggcttc aagcagcgtt aaattaatgg 780
gtaatgtgtg gatgggccgt ttgcaatacg tgggagcgta tttggcccct tcatacagca 840
cgataaacac ttcaaaagta acaggggaag tgaattttaa ccacctcact gttggcgata 900
aaaacgccgc tcaagcgggc attatcgcta ataaaaagac taatattggc acactggatt 960
tgtggcaaag cgccgggtta aacattatcg ctcctccaga aggtggctat aaggataaac 1020
ccaataatac cccttctcaa agtggtgcta aaaacgacaa aaatgaaagc gctaaaaacg 1080
acaaacaaga gagcagtcaa aataatagta acactcaggt cattaaccca cccaatagtg 1140
cgcaaaaaac agaagttcaa cccacgcaag tcattgatgg gccttttgcg ggcggcaaag 1200
acacggttgt caatatcaac cgcatcaaca ctaacgctga tggcacgatt agagtgggag 1260
ggtttaaagc ttctcttacc accaatgcgg ctcatttgca tatcggcaaa ggcggtgtca 1320
atctgtccaa tcaagcgagc gggcgctctc ttatagtgga aaatctaact gggaatatca 1380
ccgttgatgg gcctttaaga gtgaataatc aagtgggtgg ctatgctttg gcaggatcaa 1440
gcgcgaattt tgagtttaag gctggtacgg ataccaaaaa cggcacagcc acttttaata 1500
acgatattag tctgggaaga tttgtgaatt taaaggtgga tgctcataca gctaatttta 1560
aaggtattga tacgggtaat ggtggtttca acaccttaga ttttagtggc gttacagaca 1620
aagtcaatat caacaagctc attacggctt ccactaatgt ggccgttaaa aacttcaaca 1680
ttaatgaatt gattgttaaa accaatggga taagtgtggg ggaatatact cattttagcg 1740
aagatatagg cagtcaatcg cgcatcaata ccgtgcgttt ggaaactggc actaggtcac 1800
ttttctctgg gggtgttaaa tttaaaggtg gcgaaaaatt ggttatagat gagttttact 1860
atagcccttg gaattatttt gacgctagaa atattaaaaa tgttgaaatc accaataaac 1920
ttgcttttgg acctcaagga agtccttggg gcacatcaaa acttatgttc aataatctaa 1980
ccctaggtca aaatgcggtc atggattata gccaattttc aaatttaacc attcaagggg 2040
atttcatcaa caatcaaggc actatcaact atctggtccg aggtgggaaa gtggcaacct 2100
taagcgtagg caatgcagca gctatgatgt ttaataatga tatagacagc gcgaccggat 2160
tttacaaacc gctcatcaag attaacagcg ctcaagatct cattaaaaat acagaacatg 2220
ttttattgaa agcgaaaatc attggttatg gtaatgtttc tacaggtacc aatggcatta 2280
gtaatgttaa tctagaagag caattcaaag agcgcctagc cctttataac aacaataacc 2340
gcatggatac ttgtgtggtg cgaaatactg atgacattaa agcatgcggt atggctatcg 2400
gcgatcaaag catggtgaac aaccctgaca attacaagta tcttatcggt aaggcatgga 2460
aaaatatagg gatcagcaaa acagctaatg gctctaaaat ttcggtgtat tatttaggca 2520
attctacgcc tactgagaat ggtggcaata ccacaaattt acccacaaac accactagca 2580
atgcacgttc tgccaacaac gcccttgcac aaaacgctcc tttcgctcaa cctagtgcta 2640
ctcctaattt agtcgctatc aatcagcatg attttggcac tattgaaagc gtgtttgaat 2700
tggctaaccg ctctaaagat attgacacgc tttatgctaa ctcaggcgct caaggcaggg 2760
atctcttaca aaccttattg attgatagcc atgatgcggg ttatgccaga aaaatgattg 2820
atgctacaag cgctaatgaa atcaccaagc aattgaatac ggccactacc actttaaaca 2880
acatagccag tttagagcat aaaaccagcg gcttacaaac tttgagcttg agtaatgcga 2940
tgattttaaa ttctcgttta gtcaatctct ccaggagaca caccaaccat attgactcgt 3000
tcgccaaacg cttacaagct ttaaaagacc aaaaattcgc ttctttagaa agcgcggcag 3060
aagtgttgta tcaatttgcc cctaaatatg aaaaacctac caatgtttgg gctaacgcta 3120
ttgggggaac gagcttgaat aatggctcta acgcttcatt gtatggcaca agcgcgggcg 3180

tagacgctta ccttaacggg caagtggaag ccattgtggg cggttttgga agctatggtt 3240
atagctcttt taataatcgt gcgaactccc ttaactctgg ggccaataac actaattttg 3300
gcgtgtatag ccgtattttt gccaaccagc atgaatttga ctttgaagct caaggggcac 3360
tagggagcga tcaatcaagc ttgaatttca aaagcgctct attacaagat ttgaatcaaa 3420
gctatcatta cttagcctat agcgctgcaa caagagcgag ctatggttat gacttcgcgt 3480
tttttaggaa cgctttagtg ttaaaaccaa gcgtgggtgt gagctataac catttaggtt 3540
caaccaactt taaaagcaac agcaccaatc aagtggcttt gaaaaatggc tctagcagtc 3600
agcatttatt caacgctagc gctaatgtgg aagcgcgcta ttattatggg gacacttcat 3660
acttctacat gaatgctgga gttttacaag agttcgctca tgttggctct aataacgccg 3720
cgtctttaaa caccttaaa gtgaatgccg ctcgcaaccc tttaaatacc catgccagag 3780
tgatgatggg tggggaatta aaattagcta aagaagtgtt tttgaatttg ggcgttgttt 3840
atttgcacaa tttgatttcc aatataggcc atttcgcttc caatttagga atgaggtata 3900
gtttctaaat accgctctta aacccatgct caaagcatgg gtttgaaatc ttacaaaaca 3960

<210> 3
<211> 1296
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 3

```
Met Glu Ile Gln Gln Thr His Arg Lys Ile Asn Arg Pro Leu Val Ser
  1               5                  10                  15

Leu Ala Leu Val Gly Ala Leu Val Ser Ile Thr Pro Gln Gln Ser His
             20                  25                  30

Ala Ala Phe Phe Thr Thr Val Ile Ile Pro Ala Ile Val Gly Gly Ile
         35                  40                  45

Ala Thr Gly Thr Ala Val Gly Thr Val Ser Gly Leu Leu Ser Trp Gly
     50                  55                  60

Leu Lys Gln Ala Glu Glu Ala Asn Lys Thr Pro Asp Lys Pro Asp Lys
 65                  70                  75                  80

Val Trp Arg Ile Gln Ala Gly Lys Gly Phe Asn Glu Phe Pro Asn Lys
                 85                  90                  95

Glu Tyr Asp Leu Tyr Arg Ser Leu Leu Ser Ser Lys Ile Asp Gly Gly
            100                 105                 110

Trp Asp Trp Gly Asn Ala Ala Arg His Tyr Trp Val Lys Gly Gly Gln
        115                 120                 125
```

Gln Asn Lys Leu Glu Val Asp Met Lys Asp Ala Val Gly Thr Tyr Thr
130 135 140

Leu Ser Gly Leu Arg Asn Phe Thr Gly Gly Asp Leu Asp Val Asn Met
145 150 155 160

Gln Lys Ala Thr Leu Arg Leu Gly Gln Phe Asn Gly Asn Ser Phe Thr
165 170 175

Ser Tyr Lys Asp Ser Ala Asp Arg Thr Thr Arg Val Asp Phe Asn Ala
180 185 190

Lys Asn Ile Ser Ile Asp Asn Phe Val Glu Ile Asn Asn Arg Val Gly
195 200 205

Ser Gly Ala Gly Arg Lys Ala Ser Ser Thr Val Leu Thr Leu Gln Ala
210 215 220

Ser Glu Gly Ile Thr Ser Asp Lys Asn Ala Glu Ile Ser Leu Tyr Asp
225 230 235 240

Gly Ala Thr Leu Asn Leu Ala Ser Ser Ser Val Lys Leu Met Gly Asn
245 250 255

Val Trp Met Gly Arg Leu Gln Tyr Val Gly Ala Tyr Leu Ala Pro Ser
260 265 270

Tyr Ser Thr Ile Asn Thr Ser Lys Val Thr Gly Glu Val Asn Phe Asn
275 280 285

His Leu Thr Val Gly Asp Lys Asn Ala Ala Gln Ala Gly Ile Ile Ala
290 295 300

Asn Lys Lys Thr Asn Ile Gly Thr Leu Asp Leu Trp Gln Ser Ala Gly
305 310 315 320

Leu Asn Ile Ile Ala Pro Pro Glu Gly Gly Tyr Lys Asp Lys Pro Asn
325 330 335

Asn Thr Pro Ser Gln Ser Gly Ala Lys Asn Asp Lys Asn Glu Ser Ala
340 345 350

Lys Asn Asp Lys Gln Glu Ser Ser Gln Asn Asn Ser Asn Thr Gln Val
355 360 365

Ile Asn Pro Pro Asn Ser Ala Gln Lys Thr Glu Val Gln Pro Thr Gln
370 375 380

```
Val Ile Asp Gly Pro Phe Ala Gly Gly Lys Asp Thr Val Val Asn Ile
385                 390                 395                 400

Asn Arg Ile Asn Thr Asn Ala Asp Gly Thr Ile Arg Val Gly Gly Phe
                405                 410                 415

Lys Ala Ser Leu Thr Thr Asn Ala Ala His Leu His Ile Gly Lys Gly
            420                 425                 430

Gly Val Asn Leu Ser Asn Gln Ala Ser Gly Arg Ser Leu Ile Val Glu
        435                 440                 445

Asn Leu Thr Gly Asn Ile Thr Val Asp Gly Pro Leu Arg Val Asn Asn
    450                 455                 460

Gln Val Gly Gly Tyr Ala Leu Ala Gly Ser Ser Ala Asn Phe Glu Phe
465                 470                 475                 480

Lys Ala Gly Thr Asp Thr Lys Asn Gly Thr Ala Thr Phe Asn Asn Asp
                485                 490                 495

Ile Ser Leu Gly Arg Phe Val Asn Leu Lys Val Asp Ala His Thr Ala
            500                 505                 510

Asn Phe Lys Gly Ile Asp Thr Gly Asn Gly Gly Phe Asn Thr Leu Asp
        515                 520                 525

Phe Ser Gly Val Thr Asp Lys Val Asn Ile Asn Lys Leu Ile Thr Ala
    530                 535                 540

Ser Thr Asn Val Ala Val Lys Asn Phe Asn Ile Asn Glu Leu Ile Val
545                 550                 555                 560

Lys Thr Asn Gly Ile Ser Val Gly Glu Tyr Thr His Phe Ser Glu Asp
                565                 570                 575

Ile Gly Ser Gln Ser Arg Ile Asn Thr Val Arg Leu Glu Thr Gly Thr
            580                 585                 590

Arg Ser Leu Phe Ser Gly Gly Val Lys Phe Lys Gly Gly Glu Lys Leu
        595                 600                 605

Val Ile Asp Glu Phe Tyr Tyr Ser Pro Trp Asn Tyr Phe Asp Ala Arg
    610                 615                 620

Asn Ile Lys Asn Val Glu Ile Thr Asn Lys Leu Ala Phe Gly Pro Gln
625                 630                 635                 640
```

Gly Ser Pro Trp Gly Thr Ser Lys Leu Met Phe Asn Asn Leu Thr Leu
645 650 655

Gly Gln Asn Ala Val Met Asp Tyr Ser Gln Phe Ser Asn Leu Thr Ile
660 665 670

Gln Gly Asp Phe Ile Asn Asn Gln Gly Thr Ile Asn Tyr Leu Val Arg
675 680 685

Gly Gly Lys Val Ala Thr Leu Ser Val Gly Asn Ala Ala Ala Met Met
690 695 700

Phe Asn Asn Asp Ile Asp Ser Ala Thr Gly Phe Tyr Lys Pro Leu Ile
705 710 715 720

Lys Ile Asn Ser Ala Gln Asp Leu Ile Lys Asn Thr Glu His Val Leu
725 730 735

Leu Lys Ala Lys Ile Ile Gly Tyr Gly Asn Val Ser Thr Gly Thr Asn
740 745 750

Gly Ile Ser Asn Val Asn Leu Glu Glu Gln Phe Lys Glu Arg Leu Ala
755 760 765

Leu Tyr Asn Asn Asn Asn Arg Met Asp Thr Cys Val Val Arg Asn Thr
770 775 780

Asp Asp Ile Lys Ala Cys Gly Met Ala Ile Gly Asp Gln Ser Met Val
785 790 795 800

Asn Asn Pro Asp Asn Tyr Lys Tyr Leu Ile Gly Lys Ala Trp Lys Asn
805 810 815

Ile Gly Ile Ser Lys Thr Ala Asn Gly Ser Lys Ile Ser Val Tyr Tyr
820 825 830

Leu Gly Asn Ser Thr Pro Thr Glu Asn Gly Gly Asn Thr Thr Asn Leu
835 840 845

Pro Thr Asn Thr Thr Ser Asn Ala Arg Ser Ala Asn Asn Ala Leu Ala
850 855 860

Gln Asn Ala Pro Phe Ala Gln Pro Ser Ala Thr Pro Asn Leu Val Ala
865 870 875 880

Ile Asn Gln His Asp Phe Gly Thr Ile Glu Ser Val Phe Glu Leu Ala
885 890 895

```
Asn Arg Ser Lys Asp Ile Asp Thr Leu Tyr Ala Asn Ser Gly Ala Gln
            900                 905                 910

Gly Arg Asp Leu Leu Gln Thr Leu Leu Ile Asp Ser His Asp Ala Gly
        915                 920                 925

Tyr Ala Arg Lys Met Ile Asp Ala Thr Ser Ala Asn Glu Ile Thr Lys
    930                 935                 940

Gln Leu Asn Thr Ala Thr Thr Thr Leu Asn Asn Ile Ala Ser Leu Glu
945                 950                 955                 960

His Lys Thr Ser Gly Leu Gln Thr Leu Ser Leu Ser Asn Ala Met Ile
                965                 970                 975

Leu Asn Ser Arg Leu Val Asn Leu Ser Arg Arg His Thr Asn His Ile
            980                 985                 990

Asp Ser Phe Ala Lys Arg Leu Gln Ala Leu Lys Asp Gln Lys Phe Ala
        995                1000                1005

Ser Leu Glu Ser Ala Ala Glu Val Leu Tyr Gln Phe Ala Pro Lys Tyr
   1010                1015                1020

Glu Lys Pro Thr Asn Val Trp Ala Asn Ala Ile Gly Gly Thr Ser Leu
1025                1030                1035                1040

Asn Asn Gly Ser Asn Ala Ser Leu Tyr Gly Thr Ser Ala Gly Val Asp
               1045                1050                1055

Ala Tyr Leu Asn Gly Gln Val Glu Ala Ile Val Gly Gly Phe Gly Ser
           1060                1065                1070

Tyr Gly Tyr Ser Ser Phe Asn Asn Arg Ala Asn Ser Leu Asn Ser Gly
       1075                1080                1085

Ala Asn Asn Thr Asn Phe Gly Val Tyr Ser Arg Ile Phe Ala Asn Gln
   1090                1095                1100

His Glu Phe Asp Phe Glu Ala Gln Gly Ala Leu Gly Ser Asp Gln Ser
1105                1110                1115                1120

Ser Leu Asn Phe Lys Ser Ala Leu Leu Gln Asp Leu Asn Gln Ser Tyr
               1125                1130                1135

His Tyr Leu Ala Tyr Ser Ala Ala Thr Arg Ala Ser Tyr Gly Tyr Asp
           1140                1145                1150
```

Phe Ala Phe Phe Arg Asn Ala Leu Val Leu Lys Pro Ser Val Gly Val
1155 1160 1165

Ser Tyr Asn His Leu Gly Ser Thr Asn Phe Lys Ser Asn Ser Thr Asn
1170 1175 1180

Gln Val Ala Leu Lys Asn Gly Ser Ser Ser Gln His Leu Phe Asn Ala
1185 1190 1195 1200

Ser Ala Asn Val Glu Ala Arg Tyr Tyr Tyr Gly Asp Thr Ser Tyr Phe
1205 1210 1215

Tyr Met Asn Ala Gly Val Leu Gln Glu Phe Ala His Val Gly Ser Asn
1220 1225 1230

Asn Ala Ala Ser Leu Asn Thr Phe Lys Val Asn Ala Ala Arg Asn Pro
1235 1240 1245

Leu Asn Thr His Ala Arg Val Met Met Gly Gly Glu Leu Lys Leu Ala
1250 1255 1260

Lys Glu Val Phe Leu Asn Leu Gly Val Val Tyr Leu His Asn Leu Ile
1265 1270 1275 1280

Ser Asn Ile Gly His Phe Ala Ser Asn Leu Gly Met Arg Tyr Ser Phe
1285 1290 1295

<210> 4
<211> 5925
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 4
ctccatttta agcaactcca tagaccacta aagaaacttt ttttgaggct atctttgaaa 60
atctgtccta ttgatttgtt ttccattttg tttcccatgt ggatcttgtg gatcacaaac 120
gcttaattat acatgctata gtaagcatga cacacaaacc aaactatttt tagaacgctt 180
catgtgctca ccttgactaa ccatttctcc aaccatactt tagcgttgca tttgatttct 240
tcaaaaagat tcatttctta tttcttgttc ttattaaagt tctttcattt tagcaaattt 300
ttgttaattg tgggtaaaaa tgtgaatcgt cctagccttt agacgcctgc aacgatcggg 360

ctttttttcaa tattaataat gattaatgaa aaaaaaaaaa aatgcttgat attgttgtat 420
aatgagaatg ttcaaagaca tgaattgact actcaagcgt gtagcgattt ttagcagtct 480
ttgacactaa caagataccg ataggtatga aactaggtat agtaaggaga aacaatgact 540
aacgaaacca ttgaccaaca accacaaacc gaagcggctt ttaacccgca gcaatttatc 600
aataatcttc aagtagcttt tcttaaagtt gataacgctg tcgcttcata cgatcctgat 660
caaaaaccaa tcgttgataa gaacgatagg gataacaggc aagcttttga aggaatctcg 720
caattaaggg aagaatactc caataaagcg atcaaaaatc ctaccaaaaa gaatcagtat 780
ttttcagact ttatcaataa gagcaatgat ttaatcaaca aagacaatct cattgatgta 840
gaatcttcca caaagagctt tcagaaattt ggggatcagc gttaccgaat tttcacaagt 900
tgggtgtccc atcaaaacga tccgtctaaa atcaacaccc gatcgatccg aaattttatg 960
gaaaatatca tacaaccccc tatccttgat gataaagaga aagcggagtt tttgaaatct 1020
gccaaacaat cttttgcagg aatcattata gggaatcaaa tccgaacgga tcaaaagttc 1080
atgggcgtgt ttgatgagtc cttgaaagaa aggcaagaag cagaaaaaaa tggagagcct 1140
actggtgggg attggttgga tatttttctc tcatttatat ttgacaaaaa acaatcttct 1200
gatgtcaaag aagcaatcaa tcaagaacca gttccccatg tccaaccaga tatagccact 1260
accaccaccg acatacaagg cttaccgcct gaagctagag atttacttga tgaaaggggt 1320
aatttttcta aattcactct tggcgatatg gaaatgttag atgttgaggg agtcgctgac 1380
attgatccca attacaagtt caatcaatta ttgattcaca ataacgctct gtcttctgtg 1440
ttaatgggga gtcataatgg catagaacct gaaaaagttt cattgttgta tgggggcaat 1500
ggtggtcctg gagctaggca tgattggaac gccaccgttg gttataaaga ccaacaaggc 1560
aacaatgtgg ctacaataat taatgtgcat atgaaaaacg gcagtggctt agtcatagca 1620
ggtggtgaga aagggattaa caaccctagt ttttatctct acaaagaaga ccaactcaca 1680
ggctcacaac gagcattaag tcaagaagag atccaaaaca aaatagattt catggaattt 1740
cttgcacaaa ataatgctaa attagacaac ttgagcgaga aagagaagga aaaattccga 1800
actgagatta aagatttcca aaaagactct aaggcttatt tagacgccct agggaatgat 1860
cgtattgctt ttgtttctaa aaaagacaca aaacattcag ctttaattac tgagtttggt 1920
aatggggatt tgagctacac tctcaaagat tatgggaaaa aagcagataa agctttagat 1980
agggagaaaa atgttactct tcaaggtagc ctaaaacatg atggcgtgat gtttgttgat 2040
tattctaatt tcaaatacac caacgcctcc aagaatccca ataagggtgt aggcgttacg 2100
aatggcgttt cccatttaga agtaggcttt aacaaggtag ctatctttaa tttgcctgat 2160
ttaaataatc tcgctatcac tagtttcgta aggcggaatt tagaggataa actaaccact 2220
aaaggattgt ccccacaaga agctaataag cttatcaaag atttttgag cagcaacaaa 2280
gaattggttg gaaaaacttt aaacttcaat aaagctgtag ctgacgctaa aaacacaggc 2340
aattatgatg aagtgaaaaa agctcagaaa gatcttgaaa aatctctaag gaaacgagag 2400
catttagaga aagaagtaga gaaaaaattg gagagcaaaa gcggcaacaa aaataaaatg 2460
gaagcaaaag ctcaagctaa cagccaaaaa gatgagattt ttgcgttgat caataaagag 2520
gctaatagag acgcaagagc aatcgcttac gctcagaatc ttaaaggcat caaaagggaa 2580
ttgtctgata aacttgaaaa tgtcaacaag aatttgaaag actttgataa atcttttgat 2640
gaattcaaaa atggcaaaaa taaggatttc agcaaggcag aagaaacact aaaagccctt 2700
aaaggttcgg tgaaagattt aggtatcaat ccagaatgga tttcaaaagt tgaaaacctt 2760
aatgcagctt tgaatgaatt caaaaatggc aaaaataagg atttcagcaa ggtaacgcaa 2820
gcaaaaagcg accttgaaaa ttccgttaaa gatgtgatca tcaatcaaaa ggtaacggat 2880
aaagttgata atctcaatca agcggtatca gtggctaaag caacgggtga tttcagtagg 2940
gtagagcaag cgttagccga tctcaaaaat ttctcaaagg agcaattggc ccaacaagct 3000
caaaaaaatg aaagtctcaa tgctagaaaa aaatctgaaa tatatcaatc cgttaagaat 3060
ggtgtgaatg gaaccctagt cggtaatggg ttatctcaag cagaagccac aactctttct 3120
aaaaactttt cggacatcaa gaaagagttg aatgcaaaac ttggaaattt caataacaat 3180
aacaataatg gactcaaaaa cgaacccatt tatgctaaag ttaataaaaa gaaagcaggg 3240

caagcagcta gccttgaaga acccatttac gctcaagttg ctaaaaaggt aaatgcaaaa 3300
attgaccgac tcaatcaaat agcaagtggt ttgggtgttg tagggcaagc agcgggcttc 3360
cctttgaaaa ggcatgataa agttgatgat ctcagtaagg tagggctttc aaggaatcaa 3420
gaattggctc agaaaattga caatctcaat caagcggtat cagaagctaa agcaggtttt 3480
tttggcaatc tagagcaaac gatagacaag ctcaaagatt ctacaaaaca caatcccatg 3540
aatctatggg ttgaaagtgc aaaaaaagta cctgctagtt tgtcagcgaa actagacaat 3600
tacgctacta acagccacat acgcattaat agcaatatca aaaatggagc aatcaatgaa 3660
aaagcgaccg gcatgctaac gcaaaaaaac cctgagtggc tcaagctcgt gaatgataag 3720
atagttgcgc ataatgtagg aagcgttcct ttgtcagagt atgataaaat tggcttcaac 3780
cagaagaata tgaaagatta ttctgattcg ttcaagtttt ccaccaagtt gaacaatgct 3840
gtaaaagaca ctaattctgg ctttacgcaa tttttaacca atgcatttc tacagcatct 3900
tattactgct tggcgagaga aaatgcggag catggaatca agaacgttaa tacaaaaggt 3960
ggtttccaaa aatcttaaag gattaaggaa taccaaaaac gcaaaaacca ccccttgcta 4020
aaagcgaggg gttttttaat actccttagc agaaatccca atcgtcttta gtatttggga 4080
tgaatgctac caattcatgg tatcatatcc ccatacattc gtatctagcg taggaagtgt 4140
gcaaagttac gcctttggag atatgatgtg tgagacctgt agggaatgcg ttggagctca 4200
aactctgtaa aatccctatt atagggacac agagtgagaa ccaaactctc cctacgggca 4260
acatcagcct aggaagccca atcgtcttta gcggttgggc acttcacctt aaaatatccc 4320
gacagacact aacgaaaggc tttgttcttt aaagtctgca tggatatttc ctaccccaaa 4380
aagacttaac cctttgctta aaattaagtt tgattgtgct agtgggttcg tgctatagtg 4440
cgaaaattaa ttaagggtta taaagagagc ataaactaga aaaaacaagt agctataaca 4500
aagatcaagt tcaaaaaatc atagagcttt tagagcaaat tgatcgcgct cttaaccaaa 4560
gaaaaatcag aaaaaccata ggaattatca caccttataa tgcccaaaaa agacgcttgc 4620
gatcagaagt ggaaaaatac ggcttcaaga attttgatga gctcaaaata gacactgtgg 4680
atgcctttca aggtgaagag gcagatatta ttatttattc caccgtgaaa acttgtggta 4740
atctttcttt cttgctagat tctaaacgct tgaatgtggc tatttctagg gcaaaagaaa 4800
atctcatttt tgtgggtaaa aagtctttct ttgagaattt atgaagcgat gagaagaata 4860
tctttagcgc tattttgcaa gtctgtagat aggtaatctt ttccaaagat aatcattaga 4920
cattcttcgc ttcaaaacgc tttcataaat ctctctaaag cgctttataa tcaacacaat 4980
acccttatag tgtgagctat agccccttt tgggaattga gttattttga ctttaaattt 5040
ttattagcgt tacaatttga gccattcttt agcttgtttt tctagccaga tcacatcgcc 5100
gctcgcatga aattccactt tagggaatgc gtgtgcattt tttttaaggg cgtatttttg 5160
ctgcaaatat cctacaatag catcgcccga atggatgagt agggggggtg ttgaaagggc 5220
aaaatgctcc ataaaatagc cctcaatttt ttgagcgatt aagggaaaat gcgtgcaacc 5280
taaaataatc acttcgggaa aatctttaag ggagtgaaat aataacgcat gcaagtttct 5340
aacaattcgc cctctaaaat actttcttca atcaaaggca caaaaagaga agtggctaaa 5400
tgcgaaacat tcaaatagcc ttgttgtttc agggcattgt cataagcgtt ggattggatc 5460
gtcgcttttg tccctagcac taaaataggg gcgtttttat cttttacttg tcgcttgatc 5520
gctaaaatgc ttggctcaat cacgcccaca atagggattt tggaatgctt ttgcatctct 5580
tctaaagcta gagcgctcgc tgtgttgcat gccacaatca ataattcaat ctggtgcggt 5640
ttgaaaaaat ccaaagcctc taagccaaat tgcttgatcg tagtggggtc tttagtgcca 5700
taaggcactc tagccgtatc gccataatag atgatttcat caaataattg cgcttttaaa 5760
aggcttttta aaacgctaaa ccctcccaca ccgctatcaa aaacgcctat tttcatgaca 5820
ctttttttaat ttaatgggat taattaggga ttttattttt cattcattaa gtttaaaaat 5880
tcttcattgt ccttagtttg ttgcatttta gaatagacaa agctt 5925

<210> 5

<211> 1147
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 5

```
Met Thr Asn Glu Thr Ile Asp Gln Gln Pro Gln Thr Glu Ala Ala Phe
  1               5                  10                  15

Asn Pro Gln Gln Phe Ile Asn Asn Leu Gln Val Ala Phe Leu Lys Val
             20                  25                  30

Asp Asn Ala Val Ala Ser Tyr Asp Pro Asp Gln Lys Pro Ile Val Asp
         35                  40                  45

Lys Asn Asp Arg Asp Asn Arg Gln Ala Phe Glu Gly Ile Ser Gln Leu
     50                  55                  60

Arg Glu Glu Tyr Ser Asn Lys Ala Ile Lys Asn Pro Thr Lys Lys Asn
 65                  70                  75                  80

Gln Tyr Phe Ser Asp Phe Ile Asn Lys Ser Asn Asp Leu Ile Asn Lys
                 85                  90                  95

Asp Asn Leu Ile Asp Val Glu Ser Ser Thr Lys Ser Phe Gln Lys Phe
            100                 105                 110

Gly Asp Gln Arg Tyr Arg Ile Phe Thr Ser Trp Val Ser His Gln Asn
        115                 120                 125

Asp Pro Ser Lys Ile Asn Thr Arg Ser Ile Arg Asn Phe Met Glu Asn
    130                 135                 140

Ile Ile Gln Pro Pro Ile Leu Asp Asp Lys Glu Lys Ala Glu Phe Leu
145                 150                 155                 160

Lys Ser Ala Lys Gln Ser Phe Ala Gly Ile Ile Ile Gly Asn Gln Ile
                165                 170                 175

Arg Thr Asp Gln Lys Phe Met Gly Val Phe Asp Glu Ser Leu Lys Glu
            180                 185                 190

Arg Gln Glu Ala Glu Lys Asn Gly Glu Pro Thr Gly Gly Asp Trp Leu
        195                 200                 205
```

```
Asp Ile Phe Leu Ser Phe Ile Phe Asp Lys Lys Gln Ser Ser Asp Val
    210                 215                 220

Lys Glu Ala Ile Asn Gln Glu Pro Val Pro His Val Gln Pro Asp Ile
225                 230                 235                 240

Ala Thr Thr Thr Thr Asp Ile Gln Gly Leu Pro Pro Glu Ala Arg Asp
                245                 250                 255

Leu Leu Asp Glu Arg Gly Asn Phe Ser Lys Phe Thr Leu Gly Asp Met
            260                 265                 270

Glu Met Leu Asp Val Glu Gly Val Ala Asp Ile Asp Pro Asn Tyr Lys
        275                 280                 285

Phe Asn Gln Leu Leu Ile His Asn Asn Ala Leu Ser Ser Val Leu Met
    290                 295                 300

Gly Ser His Asn Gly Ile Glu Pro Glu Lys Val Ser Leu Leu Tyr Gly
305                 310                 315                 320

Gly Asn Gly Gly Pro Gly Ala Arg His Asp Trp Asn Ala Thr Val Gly
                325                 330                 335

Tyr Lys Asp Gln Gln Gly Asn Asn Val Ala Thr Ile Ile Asn Val His
            340                 345                 350

Met Lys Asn Gly Ser Gly Leu Val Ile Ala Gly Gly Glu Lys Gly Ile
        355                 360                 365

Asn Asn Pro Ser Phe Tyr Leu Tyr Lys Glu Asp Gln Leu Thr Gly Ser
    370                 375                 380

Gln Arg Ala Leu Ser Gln Glu Glu Ile Gln Asn Lys Ile Asp Phe Met
385                 390                 395                 400

Glu Phe Leu Ala Gln Asn Asn Ala Lys Leu Asp Asn Leu Ser Glu Lys
                405                 410                 415

Glu Lys Glu Lys Phe Arg Thr Glu Ile Lys Asp Phe Gln Lys Asp Ser
            420                 425                 430

Lys Ala Tyr Leu Asp Ala Leu Gly Asn Asp Arg Ile Ala Phe Val Ser
        435                 440                 445

Lys Lys Asp Thr Lys His Ser Ala Leu Ile Thr Glu Phe Gly Asn Gly
    450                 455                 460
```

Asp Leu Ser Tyr Thr Leu Lys Asp Tyr Gly Lys Lys Ala Asp Lys Ala
465 470 475 480

Leu Asp Arg Glu Lys Asn Val Thr Leu Gln Gly Ser Leu Lys His Asp
485 490 495

Gly Val Met Phe Val Asp Tyr Ser Asn Phe Lys Tyr Thr Asn Ala Ser
500 505 510

Lys Asn Pro Asn Lys Gly Val Gly Val Thr Asn Gly Val Ser His Leu
515 520 525

Glu Val Gly Phe Asn Lys Val Ala Ile Phe Asn Leu Pro Asp Leu Asn
530 535 540

Asn Leu Ala Ile Thr Ser Phe Val Arg Arg Asn Leu Glu Asp Lys Leu
545 550 555 560

Thr Thr Lys Gly Leu Ser Pro Gln Glu Ala Asn Lys Leu Ile Lys Asp
565 570 575

Phe Leu Ser Ser Asn Lys Glu Leu Val Gly Lys Thr Leu Asn Phe Asn
580 585 590

Lys Ala Val Ala Asp Ala Lys Asn Thr Gly Asn Tyr Asp Glu Val Lys
595 600 605

Lys Ala Gln Lys Asp Leu Glu Lys Ser Leu Arg Lys Arg Glu His Leu
610 615 620

Glu Lys Glu Val Glu Lys Lys Leu Glu Ser Lys Ser Gly Asn Lys Asn
625 630 635 640

Lys Met Glu Ala Lys Ala Gln Ala Asn Ser Gln Lys Asp Glu Ile Phe
645 650 655

Ala Leu Ile Asn Lys Glu Ala Asn Arg Asp Ala Arg Ala Ile Ala Tyr
660 665 670

Ala Gln Asn Leu Lys Gly Ile Lys Arg Glu Leu Ser Asp Lys Leu Glu
675 680 685

Asn Val Asn Lys Asn Leu Lys Asp Phe Asp Lys Ser Phe Asp Glu Phe
690 695 700

Lys Asn Gly Lys Asn Lys Asp Phe Ser Lys Ala Glu Glu Thr Leu Lys
705 710 715 720

Ala Leu Lys Gly Ser Val Lys Asp Leu Gly Ile Asn Pro Glu Trp Ile
725 730 735

Ser Lys Val Glu Asn Leu Asn Ala Ala Leu Asn Glu Phe Lys Asn Gly
740 745 750

Lys Asn Lys Asp Phe Ser Lys Val Thr Gln Ala Lys Ser Asp Leu Glu
755 760 765

Asn Ser Val Lys Asp Val Ile Ile Asn Gln Lys Val Thr Asp Lys Val
770 775 780

Asp Asn Leu Asn Gln Ala Val Ser Val Ala Lys Ala Thr Gly Asp Phe
785 790 795 800

Ser Arg Val Glu Gln Ala Leu Ala Asp Leu Lys Asn Phe Ser Lys Glu
805 810 815

Gln Leu Ala Gln Gln Ala Gln Lys Asn Glu Ser Leu Asn Ala Arg Lys
820 825 830

Lys Ser Glu Ile Tyr Gln Ser Val Lys Asn Gly Val Asn Gly Thr Leu
835 840 845

Val Gly Asn Gly Leu Ser Gln Ala Glu Ala Thr Thr Leu Ser Lys Asn
850 855 860

Phe Ser Asp Ile Lys Lys Glu Leu Asn Ala Lys Leu Gly Asn Phe Asn
865 870 875 880

Asn Asn Asn Asn Asn Gly Leu Lys Asn Glu Pro Ile Tyr Ala Lys Val
885 890 895

Asn Lys Lys Lys Ala Gly Gln Ala Ala Ser Leu Glu Glu Pro Ile Tyr
900 905 910

Ala Gln Val Ala Lys Lys Val Asn Ala Lys Ile Asp Arg Leu Asn Gln
915 920 925

Ile Ala Ser Gly Leu Gly Val Val Gly Gln Ala Ala Gly Phe Pro Leu
930 935 940

Lys Arg His Asp Lys Val Asp Asp Leu Ser Lys Val Gly Leu Ser Arg
945 950 955 960

Asn Gln Glu Leu Ala Gln Lys Ile Asp Asn Leu Asn Gln Ala Val Ser
965 970 975

Glu Ala Lys Ala Gly Phe Phe Gly Asn Leu Glu Gln Thr Ile Asp Lys
980 985 990

Leu Lys Asp Ser Thr Lys His Asn Pro Met Asn Leu Trp Val Glu Ser
995 1000 1005

Ala Lys Lys Val Pro Ala Ser Leu Ser Ala Lys Leu Asp Asn Tyr Ala
1010 1015 1020

Thr Asn Ser His Ile Arg Ile Asn Ser Asn Ile Lys Asn Gly Ala Ile
1025 1030 1035 1040

Asn Glu Lys Ala Thr Gly Met Leu Thr Gln Lys Asn Pro Glu Trp Leu
1045 1050 1055

Lys Leu Val Asn Asp Lys Ile Val Ala His Asn Val Gly Ser Val Pro
1060 1065 1070

Leu Ser Glu Tyr Asp Lys Ile Gly Phe Asn Gln Lys Asn Met Lys Asp
1075 1080 1085

Tyr Ser Asp Ser Phe Lys Phe Ser Thr Lys Leu Asn Asn Ala Val Lys
1090 1095 1100

Asp Thr Asn Ser Gly Phe Thr Gln Phe Leu Thr Asn Ala Phe Ser Thr
1105 1110 1115 1120

Ala Ser Tyr Tyr Cys Leu Ala Arg Glu Asn Ala Glu His Gly Ile Lys
1125 1130 1135

Asn Val Asn Thr Lys Gly Gly Phe Gln Lys Ser
1140 1145

<210> 6
<211> 546
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 6
Met Ala Lys Glu Ile Lys Phe Ser Asp Ser Ala Arg Asn Leu Leu Phe
1 5 10 15

Glu Gly Val Arg Gln Leu His Asp Ala Val Lys Val Thr Met Gly Pro

```
                20                  25                  30

Arg Gly Arg Asn Val Leu Ile Gln Lys Ser Tyr Gly Ala Pro Ser Ile
          35                  40                  45

Thr Lys Asp Gly Val Ser Val Ala Lys Glu Ile Glu Leu Ser Cys Pro
      50                  55                  60

Val Ala Asn Met Gly Ala Gln Leu Val Lys Glu Val Ala Ser Lys Thr
 65                  70                  75                  80

Ala Asp Ala Ala Gly Asp Gly Thr Thr Thr Ala Thr Val Leu Ala Tyr
                 85                  90                  95

Ser Ile Phe Lys Glu Gly Leu Arg Asn Ile Thr Ala Gly Ala Asn Pro
            100                 105                 110

Ile Glu Val Lys Arg Gly Met Asp Lys Ala Ala Glu Ala Ile Ile Asn
        115                 120                 125

Glu Leu Lys Lys Ala Ser Lys Lys Val Gly Gly Lys Glu Glu Ile Thr
    130                 135                 140

Gln Val Ala Thr Ile Ser Ala Asn Ser Asp His Asn Ile Gly Lys Leu
145                 150                 155                 160

Ile Ala Asp Ala Met Glu Lys Val Gly Lys Asp Gly Val Ile Thr Val
                165                 170                 175

Glu Glu Ala Lys Gly Ile Glu Asp Glu Leu Asp Val Val Glu Gly Met
            180                 185                 190

Gln Phe Asp Arg Gly Tyr Leu Ser Pro Tyr Phe Val Thr Asn Ala Glu
        195                 200                 205

Lys Met Thr Ala Gln Leu Asp Asn Ala Tyr Ile Leu Leu Thr Asp Lys
    210                 215                 220

Lys Ile Ser Ser Met Lys Asp Ile Leu Pro Leu Leu Glu Lys Thr Met
225                 230                 235                 240

Lys Glu Gly Lys Pro Leu Leu Ile Ile Ala Glu Asp Ile Glu Gly Glu
                245                 250                 255

Ala Leu Thr Thr Leu Val Val Asn Lys Leu Arg Gly Val Leu Asn Ile
            260                 265                 270

Ala Ala Val Lys Ala Pro Gly Phe Gly Asp Arg Arg Lys Glu Met Leu
```

```
        275                 280                 285

Lys Asp Ile Ala Ile Leu Thr Gly Gly Gln Val Ile Ser Glu Glu Leu
    290                 295                 300

Gly Leu Ser Leu Glu Asn Ala Glu Val Glu Phe Leu Gly Lys Ala Gly
305                 310                 315                 320

Arg Ile Val Ile Asp Lys Asp Asn Thr Thr Ile Val Asp Gly Lys Gly
                325                 330                 335

His Ser Asp Asp Val Lys Asp Arg Val Ala Gln Ile Lys Thr Gln Ile
            340                 345                 350

Ala Ser Thr Thr Ser Asp Tyr Asp Lys Glu Lys Leu Gln Glu Arg Leu
        355                 360                 365

Ala Lys Leu Ser Gly Gly Val Ala Val Ile Lys Val Gly Ala Ala Ser
    370                 375                 380

Glu Val Glu Met Lys Glu Lys Lys Asp Arg Val Asp Asp Ala Leu Ser
385                 390                 395                 400

Ala Thr Lys Ala Ala Val Glu Glu Gly Ile Val Ile Gly Gly Gly Ala
                405                 410                 415

Ala Leu Ile Arg Ala Ala Gln Lys Val His Leu Asn Leu His Asp Asp
            420                 425                 430

Glu Lys Val Gly Tyr Glu Ile Ile Met Arg Ala Ile Lys Ala Pro Leu
        435                 440                 445

Ala Gln Ile Ala Ile Asn Ala Gly Tyr Asp Gly Gly Val Val Val Asn
    450                 455                 460

Glu Val Glu Lys His Glu Gly His Phe Gly Phe Asn Ala Ser Asn Gly
465                 470                 475                 480

Lys Tyr Val Asp Met Phe Lys Glu Gly Ile Ile Asp Pro Leu Lys Val
                485                 490                 495

Glu Arg Ile Ala Leu Gln Asn Ala Val Ser Val Ser Ser Leu Leu Leu
            500                 505                 510

Thr Thr Glu Ala Thr Val His Glu Ile Lys Glu Glu Lys Ala Thr Pro
        515                 520                 525

Ala Met Pro Asp Met Gly Gly Met Gly Gly Met Gly Gly Met Gly Gly
```

530 535 540

Met Met
545

<210> 7
<211> 1838
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 7
aagcttgctg tcatgatcac aaaaaacact aaaaaacatt attattaagg atacaaaatg 60
gcaaaagaaa tcaaattttc agatagtgcg agaaaccttt tatttgaagg cgtgaggcaa 120
ctccatgacg ctgtcaaagt aaccatgggg ccaagaggca ggaatgtatt gatccaaaaa 180
agctatggcg ctccaagcat caccaaagac ggcgtgagcg tggctaaaga gattgaatta 240
agttgcccag tagctaacat gggcgctcaa ctcgttaaag aagtagcgag caaaaccgct 300
gatgctgccg gcgatggcac gaccacagcg accgtgctag cttatagcat ttttaaagaa 360
ggtttgagga atatcacggc tggggctaac cctattgaag tgaaacgagg catggataaa 420
gctgctgaag cgatcattaa tgagcttaaa aaagcgagca aaaaagtagg cggtaaagaa 480
gaaatcaccc aagtggcgac catttctgca aactccgatc acaatatcgg gaaactcatc 540
gctgacgcta tggaaaaagt gggtaaagac ggcgtgatca ccgttgagga agctaagggc 600
attgaagatg aattggatgt cgtagaaggc atgcaatttg atagaggcta cctctcccct 660
tattttgtaa cgaacgctga gaaaatgacc gctcaattgg ataatgctta catcctttta 720
acggataaaa aaatctctag catgaaagac attctcccgc tactagaaaa aaccatgaaa 780
gagggcaaac cgcttttaat catcgctgaa gacattgagg gcgaagcttt aacgactcta 840
gtggtgaata aattaagagg cgtgttgaat atcgcagcgg ttaaagctcc aggctttggg 900
gacagaagaa aagaaatgct caaagacatc gctattttaa ccggcggtca agtcattagc 960
gaagaattgg gcttgagtct agaaaacgct gaagtggagt ttttaggcaa agctggaagg 1020
attgtgattg acaaagacaa caccacgatc gtagatggca aaggccatag cgatgatgtt 1080
aaagacagag tcgcgcagat caaaacccaa attgcaagta cgacaagcga ttatgacaaa 1140
gaaaaattgc aagaaagatt ggctaaactc tctggcggtg tggctgtgat taaagtgggc 1200
gctgcgagtg aagtggaaat gaaagagaaa aaagaccggg tggatgacgc gttgagcgcg 1260
actaaagcgg cggttgaaga aggcattgtg attggtggcg gtgcggctct cattcgcgcg 1320
gctcaaaaag tgcatttgaa tttgcacgat gatgaaaaag tgggctatga aatcatcatg 1380
cgcgccatta aagccccatt agctcaaatc gctatcaacg ctggttatga tggcggtgtg 1440
gtcgtgaatg aagtagaaaa acacgaaggg cattttggtt ttaacgctag caatggcaag 1500
tatgtggata tgtttaaaga aggcattatt gaccccttaa aagtagaaag gatcgctcta 1560
caaaatgcgg tttcggtttc aagcctgctt ttaaccacag aagccaccgt gcatgaaatc 1620
aaagaagaaa aagcgactcc ggcaatgcct gatatgggtg gcatgggcgg tatgggaggc 1680
atgggcggca tgatgtaagc ccgcttgctt tttagtataa tctgctttta aaatcccttc 1740
tctaaatccc ccccttttcta aaatctcttt tttgggggggg tgctttgata aaaccgctcg 1800
cttgtaaaaa catgcaacaa aaaatctctg ttaagctt 1838

<210> 8
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Artificial Sequence

<400> 8
gactcgagtc gacatcga 18